



**PROSIDING
SEMINAR NASIONAL
KE 7 Tahun 2012**

Rekayasa Teknologi Industri dan Informasi
Inovasi Teknologi dan Informasi untuk
Optimalisasi Energi

**SEKOLAH TINGGI TEKNOLOGI NASIONAL
YOGYAKARTA**

SUSUNAN PANITIA

Penanggung Jawab	: Ketua STTNAS
Pengarah	: Pembantu Ketua
KetuaPelaksana	: Ir. Harianto, MT.
Sekretaris Pelaksana	: Ir. Eka Yawara, MT.
Staff Sekretariat	: 1. Sri Harjanti 2. Sunah
BendaharaPelaksana	: Drs. Sukapdi
SeksiMakalah	:
Koordinator	: Dr. Hill. Gendoet Hartono, ST., MT.
Teknik Mesin	: Dr. Ratna Kartikasari, ST, MT.
Teknik Elektro	: Tugino, ST, MT.
Teknik Sipil	: Drs. H. Triwuryanto, MT.
Teknik Geologi	: Dr. Ir. Ev. Budiadi, MS.
Teknik PWK	: Drs. Achmad Wismoro, ST, MT.
Teknik Pertambangan	: Ir. Ag. Isjudarto, MT.
Seksi Proseeding	: 1. Ir. Muhammad Abdulkadir, MT. 2. Djoko Purwanto, ST.
Seksi Acara	: Sigit Budi Hartono, ST, MT.
Seksi Publikasi, Dokumentasi	: 1. ArisWarsita, ST, MT. 2. Ferry Okto Satriya, ST. 3. Ign. Purwanto 4. H. Andiyanto, Amd.
Sponsor	: 1. Ir. Nizam Effendi 2. Sulaiman Tampubolon, ST.

DAFTAR ISI

SUSUNAN PANITIA	ii
SAMBUTAN KETUA PANITIA ReTII KE 7	iii
SAMBUTAN KETUA STTNAS	iv
DAFTAR ISI	v

TEKNIK ELEKTRO

1. Penggunaan Algoritma Differential Evolution Dalam Penyelesaian Kombinasi Pembebanan Optimal Ekonomis Dan Emisi Pada Pembangkit Listrik Termal <i>Afner Saut Sinaga</i>	1
2. Kendali Level Kecepatan Motor DC Lima Tingkat dengan Rheostat (Resistance Control) Terintegrasi Safety Deadman Pedal Pada Sistem Kereta Api Berbasis PLC (Programmable Logic Control) <i>Arifin Wibisono, Jefri Setiawan, Leonardus Heru Pratomo</i>	7
3. Pengaruh Trafik Paket Aplikasi terhadap Kinerja Jaringan dengan Manajemen Bandwidth Fifo pada Warnet Rush Yogyakarta <i>Ayu Budi Setyawati, Damar Widjaja</i>	11
4. Pengembangan Indoor Location Based Service Menggunakan Wireless Positioning pada Android <i>Dwijayanto Gusti Parrangan, Y. Sigit Purnomo Wuryo Putro, B. Yudi Dwiandiyanta</i>	17
5. Power Monitoring Berbasis Mikrokontroler <i>Freddy Kurniawan</i>	23
6. Sistem Pemerolehan Informasi Makalah Ilmiah Berbahasa Indonesia Menggunakan Struktur Data Inverted Index Berbasis Ordbms Dengan Metode Pembobotan Tf-Idf <i>Justina S. Wulandari, JB Budi Darmawan</i>	29
7. Kendali Buck-Boost Mppt Berbasis Digital <i>Matias Chosta Agryatma, Slamet Riyadi, F. Budi Setiawan</i>	35
8. Sistem Penjejak Lokasi Sumber Suara Menggunakan Interaural Time Difference <i>Muhammad Afridon, Djoko Purwanto</i>	39
9. Sistem Pemerolehan Informasi Dokumen Makalah Ilmiah Berbahasa Indonesia Menggunakan Struktur Data Inverted Index Berbasis Hash Table Dan Ordered Linkedlist <i>Reza M. Darojad, JB Budi Darmawan</i>	45
10. Desain Kontroler Fuzzy Logic untuk Robot Pembersih Sampah dalam Ruang <i>Tri Hendrawan Budianto, Irwan Dinata</i>	51
11. Kombinasi Vb dan Matlab untuk Pemrosesan Sinyal Radar Ransponder Rocket <i>Wahyu Widada</i>	57
12. Optimasi Kerja Baterai Charge-Discharge pada Sistem Pengaturan Beban (Power Management) di BTS (Base Transceiver Station) Remote Area Menggunakan Pengaturan Beban Dinamis <i>Widjonarko</i>	61
13. Perancangan Konverter Energi Berbasis Buck Chopper Untuk Panel Surya <i>Y. L. Christanto Wibowo, Ign Slamet Riyadi</i>	69
14. Desain Sistem Pembangkit Listrik Tenaga Surya Beroperasi Stand Alone dengan Konverter Ky dan Maximum Power Point Tracking Berbasis Algoritma Neuro-Fuzzy <i>Adi Kurniawan, Mochamad Ashari, Dedet C. Riawan, Ilham Pakaya</i>	75
15. Rancang Bangun Water-Meter Digital dengan Transfer Data Melalui Short Message Service (SMS) <i>Joko Prasajo, Arif Basuki, Armansyah</i>	81
16. Peningkatan Kualitas Citra Digital Dengan Metode Non-Linear Filter <i>Agus Basukesti</i>	87
17. Estimasi Kanal MIMO OFDM Berdasarkan Perubahan Nilai Signal to Noise Ratio (SNR) <i>Anggun Fitriani Isnawati</i>	93

18. Pengembangan Robot Pengikut Garis Berbasis Logika Fuzzy <i>Aji Joko Budi Pramono</i>	101
19. Perancangan Boost Konverter Sebagai Interface Antara Panel Surya Dan Beban <i>Fx Anton Yk Slamet Riyadi</i>	107
20. Pengaruh Berbagai Ekstraksi Ciri Terhadap Tingkat Pengenalan Isyarat Tutar pada Sistem Pengenalan Tutar Model Markov Tersembunyi <i>Asniar Aliyu</i>	113
21. Brushless Direct Current (BLDC) Motor Controller Using Digital Logic For Electric Vehicle <i>Bambang Sujanarko</i>	121
22. Desain dan Implementasi Maksimal Power Point Tracker dengan Kendali Tenganan untuk Sistem Pengisi Baterai <i>Banar Arianto , Leonardus. H. Pratomo</i>	125
23. Analisa Ekonomi Pemasangan Distributed Generation PLTU Prafi II Manokwari <i>Elias K. Bawan, Pandung Sarungallo</i>	131
24. Strategi Untuk Membantu Eksekutif Dalam Pengambilan Keputusan Dengan Menggunakan Data Warehouse Pengadaan Pelumas Pada PT. ABC <i>Evaristus Didik. M, Dewi. S, Felisia. L, Winnie. S</i>	137
25. Analisa Penggunaan Home Solar Cell untuk Efisiensi Pemakaian Energi Listrik (Implementasi Pemakaian Pada Perumahan Type 27/66) <i>Irfan Santoso, Tofik Hidayat</i>	143
26. Pengenalan Nada Pianika Menggunakan Fft Dan Korelasi <i>Dionysius Edwin Surya, Linggo Sumarno</i>	151
27. Maximum Power Point Tracking Menggunakan Artificial Neural Network Untuk Sistem PV Terhubung Grid Melalui Wide Range Input Inverter <i>Muhammad Syafei Gozali, Dedet Candra Riawan, Mochamad Ashari</i>	159
28. Rancang Bangun Alat Bantu Penentuan Lokasi Kincir Angin Pada PLT ANGIN <i>Tito Yuwono, Budi Astuti, Febrian Fariz</i>	165
29. Pengembangan E-Procurement dengan Menggunakan Kerangka Kerja Cobit (Studi Kasus : Kementerian Keuangan Timor - Leste) <i>Onorio Dos Santos, Benyamin L. Sinaga, Paulus Mudjihartono</i>	171
30. Perancangan Catu Daya Dengan High Frequency Transformator Berbasis Kendali Digital <i>Dionisius Wahyu Pradana, Ign. Slamet Riyadi</i>	177
31. Dummy Load Untuk Beban 450 Watt <i>Pernandes, Martanto</i>	181
32. Penentuan Tegangan Penyalaan (Ignition-Voltage) pada Kendaraan Tangki Pembawa Bahan Bakar Cair Premium Dan Gas Lpg dalam Kawasan Bermedan Listrik <i>Budi Utama</i>	187
33. Desain Sistem Jaring Kecil (Micro Grid System) Berbasis Photovoltaic Menggunakan Kontrol Artificial Neural Network <i>Wan Muhammad Faizal, M. Ashari, Heri Suryo.A</i>	195
34. Chopper-Inverter Sebagai Interface Pv dan Sistem Kelistrikan <i>Ricky Gondo Atmodjo, Slamet Riyadi,</i>	201
35. Perbandingan Kinerja Empat Metode Prototipe Alat Ukur Kadar Curcuminoid pada Rimpang Kunyit (Curcuma Domestica) <i>Bernadeta Wuri Harini, Rini Dwiastuti, Lucia Wiwid Wijayanti</i>	205
36. Analisis Pengaruh Penempatan UPFC Terhadap Minimisasi Rugi Saluran Transmisi <i>Petrus Setyo Prabowo</i>	211
37. Analisa dan Perancangan Portal Web Untuk Konsultan Pajak <i>Stefanus Cendra Hogi Sopacua, Flourensia Sapty Rahayu, Eduard Rusdianto</i>	217
38. Prototipe Multigain Gyroscope Untuk Aplikasi Roket <i>Sri Kliwati</i>	223
39. Rancang Bangun Robot Animaloid Berkaki Empat	

<i>Tugino, Septian Andra, Sudiana</i>	227
40. Pengaruh Tegangan Impuls terhadap Ketahanan Arester Tegangan Rendah <i>Diah Suwarti</i>	231
41. Perancangan Sistem Embedded berbasis FPGA <i>Totok Mujiono, Tasripan, Pujiono</i>	237
42. Desain Sistem Dual Inputs Sepic – Bidirectional Converter untuk Manajemen Energi Sistem Pembangkit Photovoltaic Pada Area Terpencil <i>Daniar Fahmi, Dedet C. Riawan, M. Ashari</i>	241
43. Kelayakan Penerapan Sistem Kendali Suhu Dan Kelembaban Pada Budidaya Jarum Tiram di Jogjakarta <i>Hendra Setiawan, Sholichin</i>	249

TEKNIK MESIN

1. Model Peningkatan Kualitas Layanan Kesehatan di Jawa Timur Melalui Integrasi Metode Servqual, Lean dan Six Sigma Untuk Meningkatkan Kepuasan Konsumen <i>Hana Catur Wahyuni, Wiwik Sulistiyowati</i>	255
2. Pengaruh Penambahan Krom dan Tempo terhadap Kekerasan, Keausan dan Struktur Mikro Ball Mill <i>Sumpena, Subarmono, R. Soekrisno</i>	262
3. Perancangan Dan Pembuatan Mesin Produksi Palet Ikan Dengan Pendekatan Ergonomis <i>Tofik Hidayat, Irfan Santoso</i>	269
4. Pengaruh Kadar Si Terhadap Struktur Mikro Dan Sifat Mekanik Kandidat Baja Ringan Paduan Fe-Al-Mn <i>Ratna Kartikasari, Sutrisna</i>	275
5. Studi Peningkatan Daya dan Torsi dengan Pemasangan Air Tube pada Saluran Intake Manifold Sepeda Motor Dua Langkah <i>Harjono</i>	280
6. Pengembangan Program Pendukung Keputusan Untuk Estimasi Manufacturing Cost Pada Perancangan Cold Storage Menggunakan Panel Surya <i>Boni Sena, Fauzun, Endang Suhendar</i>	284
7. Aplikasi Interferometer Michelson Untuk Pengukuran Regangan Pada Mesin Uji Tari <i>Budi Setyahandana, Martanto, Ronny Dwi Agusulistyo</i>	289
8. Efek Perubahan Ukuran Diameter Header Knalpot terhadap Konsumsi Bahan Bakar dan Akselerasi Kendaraan Pada Motor 4 Tak <i>Aji Pranoto</i>	296
9. Penentuan Jalur Terpendek Petugas Kebersihan Sampah Di Lingkungan Perumahan Dosen UGM Sekip Menggunakan Algoritma Semut <i>Andhi Akhmad Ismail, Radhian Krisnaputra</i>	302
10. Pengaruh Perubahan Debit Aliran Udara-Air Terhadap Respon Amplitudo dengan Menggunakan Gelombang Ultrasonik Pada Aliran Stratified Horizontal <i>Mukhlis, Bramantya, Hermawan</i>	308
11. Pengaruh Volume Ruang Bakar Terhadap Kinerja Mesin Pulse Jet <i>Lambertus Dwi Setiawan</i>	314
12. Studi Eksperimen Batas Mampu Bakar Campuran LPG / CO ₂ sebagai Refrigeran Alternatif <i>Nasrul Iminnafik</i>	318
13. Menyelidiki Pengaruh Pemasangan Pelat-Pelat Pengarah Angin pada Keliling Lingkaran Luar Sudu Kincir Terhadap Unjuk Kerja Tiga Model Kincir Angin Savonius <i>Rines</i>	322
14. Identifikasi, Pemodelan dan Kompensasi Ketidaktelitian Open Loop Control System Pada Mesin Milling CNC Mini <i>Ignatius Aris Hendaryanto, M. Arif Wibisono, Herianto</i>	329
15. Analisis Pengerasan Permukaan Metode Flame Hardening dengan Pencekaman Spesimen Sistem Vertikal Pada Baja S45C	

SISTEM PEMEROLEHAN INFORMASI MAKALAH ILMIAH BERBAHASA INDONESIA MENGGUNAKAN STRUKTUR DATA INVERTED INDEX BERBASIS ORDBMS DENGAN METODE PEMBOBOTAN TF-IDF

Justina S. Wulandari¹, JB Budi Darmawan²

^{1,2} Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Sanata Dharma
Kampus III Paingan Maguwohardjo Depok Sleman Yogyakarta
¹wulwulmocil@gmail.com, ²jbbudi@gmail.com

ABSTRAK

Setiap tahun makalah ilmiah berbahasa Indonesia semakin bertambah banyak sehingga diperlukan sistem pemerolehan informasi untuk mencari dokumen yang relevan. Sebagian besar sistem pemerolehan informasi dan web pencarian menerapkan *inverted index* yang terbukti efisien dalam menjawab *query*. Pada penelitian sebelumnya, penggunaan *inverted index* pada *database* menawarkan beberapa keuntungan dan implementasi *inverted index* dengan Bahasa Indonesia menggunakan ORDBMS memiliki waktu akses yang lebih baik dibandingkan menggunakan RDBMS. Pada makalah ini penulis mencoba untuk mengimplementasikan sistem pencarian makalah ilmiah berbahasa Indonesia berdasarkan seluruh isi teks dalam dokumen menggunakan struktur data *inverted index* pada ORDBMS dengan metode pembobotan TF-IDF. Percobaan ini menggunakan *corpus* 281 dokumen makalah ilmiah berbahasa Indonesia yang berisi 25737 term. Waktu yang dibutuhkan untuk proses pencarian untuk operasi AND dengan 1 sampai 2 kata kunci kurang dari 0,1 detik, 3 kata kunci kurang dari 0,3 detik, dan 4 kata kunci kurang dari 0,5 detik.

Kata kunci : *inverted index*, ORDBMS, TF-IDF, pemerolehan informasi

1. PENDAHULUAN

1.1. Latar Belakang

Kemajuan dunia pendidikan maupun riset dibidang teknologi cukuplah pesat, khususnya teknologi informasi yang dapat dilihat dari banyak bermunculan makalah penelitian. Setiap tahun makalah ilmiah akan bertambah banyak seiring dengan banyaknya orang yang melakukan penelitian. Pencari makalah pun ingin menemukan makalah yang sesuai dengan kebutuhannya. Penyaringan makalah secara manual dilakukan dengan membaca makalah yang kemudian diklasifikasikan dengan kesesuaian makalah yang diinginkan. Untuk mempermudah pencari makalah, maka diperlukan sistem pemerolehan informasi untuk mencari dokumen makalah yang relevan.

Sebagian besar sistem pemerolehan informasi dan web pencarian menerapkan *inverted index* yang terbukti efisien dalam menjawab *query* [2]. Penggunaan *inverted index* pada DBMS (*Database Management System*) menawarkan beberapa kelebihan dan kekurangannya [7]. Implementasi *inverted index* dengan Bahasa Indonesia pada ORDBMS (*Object Relational Database Management System*) memiliki waktu akses yang lebih baik dibanding menggunakan RDBMS (*Relational Database Management System*) [4]. Proses pembentukan indeks menggunakan rumus Savoy berdasarkan frekuensi kata, jumlah dokumen, dan frekuensi terbesar pada sebuah dokumen yang mampu memberikan bobot yang lebih spesifik pada dua dokumen yang mempunyai frekuensi sama, sehingga mudah untuk diperingkat [5]. Penerapan indeks ini sudah pernah digunakan pada dokumen berbahasa Indonesia dan dapat menghasilkan

dokumen *ter-retrieve* yang sesuai, dengan mengurutkan bobot dokumen-dokumen yang ditemukan [5]. Penerapan struktur data *inverted index* di ORDBMS dapat menggunakan *collection* yang struktur datanya mirip dengan struktur data *inverted index* klasik [4].

1.2. Tujuan

Tujuan dari penelitian ini adalah mengimplementasikan sistem pemerolehan informasi untuk pencarian makalah ilmiah berbahasa Indonesia dan mengamati unjuk kerja penggunaan struktur data *inverted index* berbasis ORDBMS dengan metode pembobotan TF-IDF. Penelitian ini bermanfaat sebagai alternatif penerapan *inverted index* pada ORDBMS dengan metode pembobotan TF-IDF untuk memperoleh kelebihan yang ditawarkan.

1.3. Metode Penelitian

Dalam penelitian ini dilakukan tahap-tahap sebagai berikut :

1. Studi pustaka penerapan pemerolehan informasi untuk menjawab *query* dalam ORDBMS menggunakan pembobotan TF-IDF dengan model pemerolehan *boolean* AND.
2. Pengumpulan dokumen-dokumen makalah ilmiah berbahasa Indonesia sebagai *corpus*.
3. Implementasi penerapan struktur data *inverted index* dalam ORDBMS dengan metode pembobotan TF-IDF menggunakan rumus Savoy.
4. Pengamatan unjuk kerja waktu pada operasi pencarian AND dengan dua belas kelompok kata berdasarkan jumlah dokumen yang mengandung istilah tertentu (dfk) dan jumlah kata kunci pada

query pencarian. Dua belas kelompok kata ini adalah kelompok kata yang mempunyai dfk minimum 1 sampai 2 dokumen, dfk kurang lebih 140 dokumen, dfk maksimum 281 dokumen dengan *query* pencarian dari 1 kata kunci sampai 4 kata kunci.

Penelitian ini dilakukan dengan menggunakan sebuah komputer dengan spesifikasi sebagai berikut:

a. Perangkat lunak

1. Sistem operasi, Oracle Linux Server Release 5.8
2. Oracle 11g R2 Enterprise Edition,
3. Oracle SQL Developer 3.0.04,
4. Java JDK 1.6.0 dan OJDBC
5. Netbeans 6.9.1

b. Perangkat keras

1. Prosesor: Intel XEON E5620 (4 Core, 2, 40 GHz)
2. Memori RAM: 8 GB RDIMM
3. Hardisk: RAID 5 Logical 2TB
4. Motherboard: HP Proliant ML 350 G6

2. LANDASAN TEORI

2.1. Collection pada ORDBMS

ORDBMS merupakan penerapan model berorientasi obyek pada Oracle dan perluasan dari RDBMS. Salah satu fitur pada ORDBMS adalah *collection*. *Collection* digunakan untuk menyimpan *multiple values* dalam satu kolom dari sebuah tabel yang menghasilkan *nested table* dimana sebuah kolom dalam sebuah tabel berisi tabel lain [3].

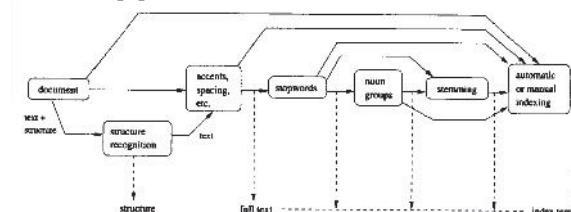
Nested table adalah *unordered set* dari elemen data yang *data typenya* sama dan memiliki sebuah kolom dari *object type*. Jika kolom adalah sebuah *object type*, tabel juga dapat dilihat sebagai tabel *multi-column*, dengan sebuah kolom untuk setiap atribut dari *object type* [3].

Pada Oracle semua data *nested table* disimpan dalam satu tabel, yang terkait dengan tabel terlampir atau *object type* [8].

2.2. Information Retrieval

Dokumen dalam koleksi sering diwakili melalui set indeks *term* atau *keyword*. *Keyword* tersebut dapat diekstraksi langsung dari teks dokumen atau ditentukan secara manual (dibuat oleh spesialis seperti banyak dilakukan pada bidang *information science*). Indeks *term* ini menyediakan suatu *logical view* dari dokumen. Komputer modern memungkinkan representasi suatu dokumen dengan menggunakan seluruh teks yang terdapat dalam dokumen tersebut. Dalam hal ini, sistem temu balik disebut mengadopsi *full text logical view* dari dokumen. Jika koleksi dokumen sangat besar, maka komputer modern akan mengurangi jumlah *set term* indeks melalui proses penghapusan *stopwords*, *stemming*, dan indentifikasi *noun groups* (menghilangkan *adjectives*, *adverbs*, dan *verbs*). Proses tersebut disebut *text operation* (operasi teks)

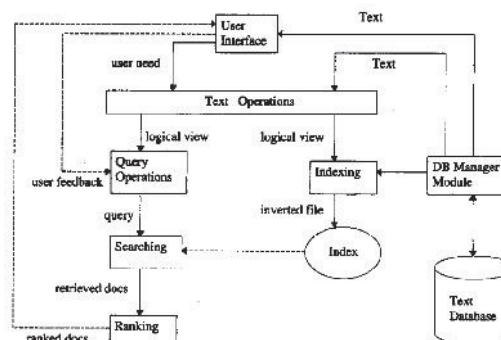
yang akan mengurangi kompleksitas dari representasi dokumen dan mengubah *logical view* dari *full text* menjadi set indeks *term*. Gambar 1 mengilustrasikan beberapa *intermediate logical view* yang akan digunakan oleh suatu sistem pemerolehan informasi [2].



Gambar 1. Logical view dari sebuah dokumen: dari *full text* menjadi sebuah set indeks *term*. [2]

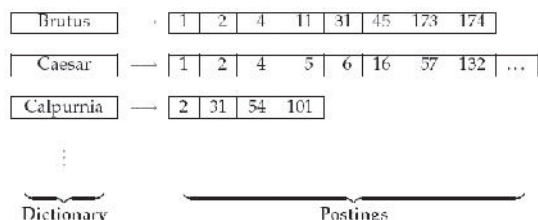
Proses *retrieval* terdapat beberapa tahap. Pertama-tama, sebelum proses *retrieval* diinisiasi, diperlukan pendefinisian *database* teks. Hal ini dilakukan dengan melakukan identifikasi terhadap dokumen-dokumen yang akan digunakan, operasi yang akan dilakukan terhadap teks, dan model teks (struktur teks dan elemen mana saja dari teks yang dapat *directly retrieve*). *Text operations* mentransformasikan dokumen asal menjadi *logical view* dokumen tersebut. Setelah *logical view* dokumen diperoleh, dibuatlah indeks *term* pada *database* untuk mempercepat proses pencarian terhadap jumlah data yang besar. Struktur indeks yang paling banyak digunakan adalah *inverted file* seperti pada Gambar 2. *Database* dokumen telah diindeks, maka proses *retrieval* dapat diinisiasi. Pengguna menentukan kebutuhannya yang kemudian ditransformasikan oleh *text operation* yang sama digunakan pada koleksi dokumen. *Query* kemudian ditransformasi untuk mendapatkan dokumen *retrieval*. Struktur indeks dibuat sebelumnya agar dapat mempengaruhi cepatnya pemrosesan *query* [2].

Sebelum dikirimkan ke pengguna, dokumen *retrieval* diperingkat dahulu berdasarkan kemungkinan relevansi. Pengguna kemudian memeriksa set dokumen peringkat untuk mendapatkan informasi yang berguna baginya [2].



Gambar 2. Proses pemerolehan informasi [2].

Representasi struktur data *inverted index* pada Gambar 3 menunjukkan *dictionary* berisi kumpulan *term* (t) yang masing-masing *term* mempunyai *postings list* (atau *inverted list*) yang berisi kumpulan dokumen (d) [6].



Gambar 3. Representasi *Inverted Index* [6]

Inverted index pada Gambar 3, dapat dilakukan operasi *boolean* dasar. Operasi AND dengan *n operand* akan melibatkan *n postings list*. Operasi Caesar AND Calpurnia akan menghasilkan dokumen 2.

2.3. Pembobotan TF-IDF

Teknik pembobotan Savoy (1993) adalah sebagai berikut [5]:

$$W_{ik} = \text{ntf}_{ik} * \text{nidf}_k,$$

$$\text{ntf}_{ik} = \frac{tf_{ik}}{\text{Max}_j tf_{ij}} \quad \text{dan} \quad \text{nidf}_k = \frac{\log \left[\frac{n}{df_k} \right]}{\log(n)}$$

Dimana :

- W_{ik} adalah bobot istilah k pada dokumen i.
- tf_{ik} merupakan frekuensi dari istilah k dalam dokumen i.
- n adalah jumlah dokumen dalam kumpulan dokumen.
- df_k adalah jumlah dokumen yang mengandung istilah k.
- $\text{Max}_j tf_{ij}$ adalah frekuensi istilah terbesar pada satu dokumen.
- W_d = bobot sebuah dokumen

Pada teknik pembobotan ini, bobot istilah dinormalisasi. Dalam menentukan bobot suatu istilah tidak hanya berdasarkan frekuensi *term*, tetapi juga berdasarkan frekuensi terbesar pada dokumen bersangkutan. Hal ini untuk menentukan posisi relatif bobot dari *term* dibanding dengan *term-term* lain di dokumen yang sama. Selain itu teknik ini juga memperhitungkan jumlah dokumen yang mengandung *term* bersangkutan dan jumlah dokumen. Hal ini untuk menentukan posisi relatif bobot *term* bersangkutan pada suatu dokumen dibandingkan dengan dokumen-dokumen lain yang memiliki *term* yang sama. Sehingga jika sebuah *term* memiliki frekuensi yang sama pada dua dokumen belum tentu memiliki bobot yang sama [5].

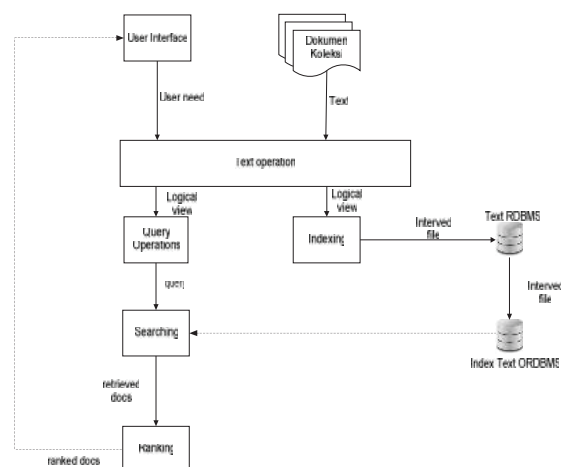
Menurut Hasibuan, pada teknik *Boolean* berperingkat dengan menggunakan pengindeksan

berdasarkan rumus Savoy, telah terjadi perbaikan dari hasil temu-kembali dimana dokumen yang di-retrieve telah diberi bobot dan diperingkat berdasarkan bobotnya yang akan memudahkan pengguna untuk memilih dokumen yang benar-benar relevan dari dokumen-dokumen hasil yang ditampilkan [5].

3. HASIL PENELITIAN DAN PEMBAHASAN

3.1. Arsitektur Sistem

Arsitektur dari sistem pemerolehan informasi yang digunakan dalam penelitian ini dapat dilihat pada Gambar 4. Kumpulan dokumen koleksi berupa *text file* dengan ekstensi pdf. Sistem akan melakukan proses *text operation* pada dokumen menggunakan algoritma *stemming* Nazief dan Adriani untuk teks Bahasa Indonesia [1], dilanjutkan proses *indexing* untuk pembentukan *inverted index* yang disimpan dalam *database text* dengan menggunakan teknologi RDBMS. Sistem mengambil kumpulan *term* hasil *indexing* dari *database text*, kemudian menyimpan dalam bentuk *inverted index* ke *database index ready* dengan teknologi ORDBMS. Sedangkan, untuk proses pencarian pengguna memasukkan kebutuhan pengguna dalam bentuk kata kunci dan sistem akan melakukan proses *text operation*, dilanjutkan dengan proses *query operations* dan *searching*, kemudian dilakukan proses *ranking* berdasarkan bobotnya yang kemudian akan ditampilkan dokumen hasil melalui *user interface*.



Gambar 4. Rancangan sistem pemerolehan informasi

3.2. Data yang Digunakan

Tabel JURNAL dan TERM yang digunakan pada percobaan ini menggunakan teknologi ORDBMS. Struktur tabel JURNAL ditampilkan pada Gambar 5. Tabel ini untuk menyimpan data dari sebuah dokumen yang terdiri dari kolom JURNAL_ID sebagai *primary key*, JUDUL untuk menyimpan judul dari sebuah dokumen, TAHUN untuk menyimpan tahun dari sebuah dokumen, NAMA_PENULIS untuk menyimpan nama penulis dari sebuah dokumen, PATH untuk menyimpan

nama file dari dokumen, MAXJTFIJ untuk menyimpan frekuensi terbesar dari sebuah dokumen.

```
CREATE or REPLACE TYPE JurnalType AS OBJECT
(jurnal_id    NUMBER,
judul        VARCHAR2(256),
tahun        NUMBER,
nama_penulis VARCHAR2(100),
Maxjtfij     NUMBER,
path         VARCHAR2(256));
/

CREATE TABLE Jurnal OF JurnalType
(jurnal_id NOT NULL,
PRIMARY KEY (jurnal_id));
```

Column Name	Data Type	Nullable
JURNAL_ID	NUMBER	No
JUDUL	VARCHAR2(256 BYTE)	Yes
TAHUN	NUMBER	Yes
NAMA_PENULIS	VARCHAR2(100 BYTE)	Yes
MAXJTFIJ	NUMBER	Yes
PATH	VARCHAR2(256 BYTE)	Yes

Gambar 5. Struktur tabel JURNAL pada ORDBMS

Untuk struktur tabel TERM ditampilkan pada Gambar 6. Tabel ini terdiri dari kolom TERM_ID sebagai *primary key*, TERM untuk mewakili kata, DFK untuk menyimpan jumlah dokumen yang mengandung term, NIDFK untuk menyimpan hasil perhitungan $(\log(\text{DFK}/N))/\log N$, POSTING_TERM bertipe *nested table* yang mempresentasikan *posting list* untuk menyimpan kolom-kolom, yaitu TFK (frekuensi *term* pada suatu dokumen), NTFIK (hasil perhitungan TFK/frekuensi terbesar pada suatu dokumen), W (bobot dari sebuah *term*), JURNAL_POSTING (kolom untuk merujuk ke objek tabel JURNAL).

```
CREATE or REPLACE TYPE PostingListType AS OBJECT
(dfk NUMBER,
ntfik DECIMAL(10,5),
w NUMBER,
jurnal_posting REF JurnalType);
/

CREATE TYPE PostingListNestedType AS TABLE OF PostingListType;
/

CREATE OR REPLACE TYPE termType AS OBJECT
(term_id    NUMBER,
term       VARCHAR2(50),
dfk        NUMBER,
nidfk      NUMBER,
posting_term PostingListNestedType);
/

CREATE TABLE term OF termType
(term_id NOT NULL,
PRIMARY KEY (term_id),
NESTED TABLE posting_term STORE AS posting_term_table);
```

Column Name	Data Type	Nullable
TERM_ID	NUMBER	No
TERM	VARCHAR2(50 BYTE)	Yes
DFK	NUMBER	Yes
NIDFK	NUMBER	Yes
POSTING_TERM	POSTINGLISTNESTEDTYPE	Yes

Gambar 6. Struktur tabel TERM pada ORDBMS

3.3. Hasil Percobaan

Dokumen hasil dari operasi AND dengan metode pembobotan TF-IDF adalah dokumen yang terdapat pada semua kata kunci yang kemudian dilakukan proses *ranking* berdasarkan jumlah nilai bobotnya sebelum ditampilkan ke pengguna.

Dari hasil pengamatan waktu akses *query* pada Tabel 1 terdapat peningkatan waktu akses untuk semua kelompok kata dfk dengan meningkatnya jumlah kata kunci. Pada kelompok dfk minimum 1 sampai 2 dokumen untuk 1 kata kunci memerlukan waktu akses *query* 0,0026 detik dan 4 kata kunci memerlukan waktu akses *query* hingga 0,0074 detik. Sedangkan, pada dfk maksimum 281 dokumen untuk 1 kata kunci memerlukan waktu akses *query* 0,013 detik dan 4 kata kunci memerlukan waktu akses *query* hingga 0,492 detik. Hal ini dikarenakan sistem akan melakukan pencarian dokumen pada *database* berdasarkan kata kunci yang dimasukkan untuk mendapatkan *posting list* dokumen. Untuk kata kunci lebih dari 1 akan dilakukan operasi AND dari kumpulan *posting list* dokumen yang didapatkan dari *database*, dilanjutkan dengan proses *ranking* berdasarkan jumlah bobotnya sebelum dokumen hasil ditampilkan ke pengguna.

Pada Gambar 7 terlihat bahwa pencarian menggunakan 2 sampai 4 kata kunci mengalami kenaikan waktu akses *query* secara linear dan cukup tajam dibandingkan dengan 1 kata kunci, hal ini disebabkan pada pencarian lebih dari 1 kata kunci akan dilakukan operasi AND yang dilanjutkan proses *ranking* dokumen hasil. Sedangkan, pada pencarian untuk 1 kata kunci hanya melakukan proses pencarian tanpa menggunakan operasi AND yang dilanjutkan proses *ranking* dokumen hasil.

Pencarian menggunakan kata kunci lebih dari 1 kata terlebih dahulu akan dilakukan operasi AND untuk 2 kata dan hasil dari setiap operasi akan dilakukan operasi AND hingga kata terakhir. Hasil dari operasi AND dengan kata terakhir akan dilakukan proses *ranking*. Maka, semakin banyaknya kata kunci akan berpengaruh ke waktu akses *query*.

Dari Tabel 1 terlihat bahwa peningkatan jumlah dfk dari 1 sampai 281 dokumen akan meningkatkan waktu akses *query*. Untuk pencarian jumlah 1 kata kunci pada dfk minimum 1 sampai 2 dokumen memerlukan waktu akses *query* 0,0026 detik dan dfk maksimum 281 dokumen mengalami peningkatan waktu akses *query* hingga 0,013 detik. Sedangkan, untuk jumlah 4 kata kunci pada dfk minimum 1 sampai 2 dokumen memerlukan waktu 0,0074 detik dan dfk maksimum 281 dokumen mengalami peningkatan waktu akses *query* hingga 0,492 detik.

Tabel 1. Rata-rata waktu *query* operasi AND untuk 3 kelompok dfk dengan 1 sampai 4 kata kunci

dfk	jumlah kata kunci	rata-rata waktu <i>query</i> (detik)
1-2	1	0,0026
	2	0,0038
	3	0,0057
	4	0,0074
±140	1	0,0109
	2	0,0126
	3	0,0712
	4	0,1708
281	1	0,013
	2	0,079
	3	0,2958
	4	0,492

**Gambar 7.** Grafik rata-rata waktu *query* untuk 3 kelompok dfk dengan menggunakan 1 sampai 4 kata kunci

4. PENUTUP

4.1. Kesimpulan

Dari hasil penelitian menggunakan *corpus* 281 dokumen makalah ilmiah berbahasa Indonesia yang menghasilkan hampir 25737 *term*, dapat disimpulkan penerapan *inverted index* ke dalam ORDBMS menggunakan metode pembobotan TF-IDF dengan *nested table collection* untuk operasi AND pada dfk minimum 1 sampai 2 dokumen dengan 1 kata kunci memerlukan waktu 0,0026 detik dan 4 kata kunci memerlukan waktu hingga 0,0074 detik. Sedangkan, pada dfk maksimum 281 dokumen dengan 1 kata kunci memerlukan waktu 0,013 detik dan 4 kata kunci memerlukan waktu hingga 0,492 detik. Pada pencarian menggunakan kata kunci 2 sampai 4 mengalami kenaikan secara linear, dikarenakan pada pencarian lebih dari 1 kata kunci akan dilakukan operasi AND. Sedangkan, pada pencarian menggunakan 1 kata kunci hanya dilakukan proses pencarian tanpa melakukan operasi AND. Semakin banyak kata kunci yang digunakan akan membuat semakin banyaknya operasi AND yang dilakukan.

Peningkatan jumlah dfk dan kata kunci akan meningkatkan lamanya waktu pencarian, dikarenakan sistem akan mengembalikan banyaknya *posting list* sesuai dengan jumlah kata kunci dan banyaknya dokumen dalam *posting list* sesuai dengan jumlah dfk.

Penerapan *inverted index* pada ORDBMS dengan menggunakan metode pembobotan TF-IDF menjadi salah satu alternatif untuk mendapatkan kelebihan yang ditawarkan.

4.2. Saran

Penelitian selanjutnya dapat dilakukan untuk operasi pencarian OR dan NOT atau operasi pencarian AND, OR, NOT menggunakan metode pembobotan lain.

DAFTAR PUSTAKA

- [1] Agusta, Ledy. (2009). *Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia*. Jurnal Konferensi Nasional Sistem dan Informatika 2009, Bali.
- [2] Baeza-Yates, R., Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. AddisonWesley.
- [3] Connolly, T., Begg, C., (2005). *Database Systems: A Practical Approach to Design, Implementation, and Management*, 4th edition, England, Pearson Education Limited.
- [4] Darmawan, J.B.Budi. (2012). *Inverted Index untuk Medukung Model Pemerolehan Boolean Menggunakan RDBMS VS ORDBMS*. Prosiding Konferensi Nasional Sistem Informasi 2012, Bali.
- [5] Hasibuan, Zainal A., & Andri, Yofi. (2001). *Penerapan Berbagai Teknik Sistem Temu-Kembali Informasi Berbasis Hiperteks*. Jurnal Ilmu Komputer dan Teknologi Informasi, Volume 1, Nomor 2.
- [6] Manning, Christopher, D., Raghavan, Prabhakar., Schütze, Hinrich. (2009). *An Introduction to Information Retrieval*. England: Cambridge University Press.
- [7] Papadakos P., Theoharis Y., Marketakis Y., Armenatzoglou N., Tzitzikas Y. (2008). *Mitos: Design and Evaluation of a DBMS-based Web Search Engine*. IEE.
- [8] Pardede, Eric., Rahayu, Wenny., Taniar, David. (2006). *Object Oriented OracleTM*. America: IRM Press.